

Remarks

Status of application

Claims 1-63 were examined and stand rejected in view of prior art. The claims have been amended to further clarify Applicant's invention. Reexamination and reconsideration are respectfully requested.

The invention

Applicant's invention comprises a system and methodology which allows users to capture dynamic content from a variety of sources such as Web pages, databases, and XML documents and replay or reuse the captured content. Applicant's solution provides users with an easy-to-use and flexible means to capture and aggregate data from a variety of sources. For example, data "records" extracted from one Web page may be captured and aggregated with similar data records extracted from several other Web pages. The aggregate, extracted records may then be deployed to a Web site, a database, a XML document, or the like. Applicant's solution can be used to capture data elements from one or more data sources (e.g., Web pages) and then harvest the captured information for use in a number of different applications. For example, information extracted from HTML pages and aggregated into an XML document can be integrated into a database which is deployed for use on various platforms to provide access to the collected information.

Applicant's invention makes it is possible for a user to extract information from sources available on the Internet (e.g., Web pages) in unstructured or semi-structured formats such as articles, stories, and other text on HTML pages and then to aggregate the information based on attributes specified by the user. For example, a user may use the visual content development interface of Applicant's solution to identify and tag (or "fingerprint") text-runs such as article titles, bylines, dates, and body text on a given Web page. A feature extraction tag (or object) is created based on attributes of the extracted information. This feature tag provides access to the captured information in the future, even if the Web page may have modified in the interim by the addition of new graphics or information. Applicant's invention also enables the captured information to be translated into a structured format based on a mapping specified by the user. For example, a user may organize data extracted from a web page into columns such as

"title", "author", "date" and "text". Data extracted, aggregated and structured using the present invention can be deployed to a number of portal/Web application servers and/or can be saved in an XML document, spreadsheet or database to enable it to be reused in various applications. The solution includes tools and utilities to enable these content collection and aggregation tasks to be performed by non-programmer users.

Prior art rejections

A. Section 102 rejection: Whitledge

Claims 1-63 stand rejected under 35 U.S.C. 102(e) as being anticipated by US Patent No. 6,925,595 to Whitledge et al (hereinafter "Whitledge"). The Examiner's rejection of Applicant's Claim 1 as follows is representative of the rejection of Applicant's claims as anticipated by Whitledge:

As to claim 1, Whitledge discloses a method for extracting and structuring items of data from content available via the Internet, the method comprising:

receiving input of a user (i.e. a user is able select one or more hypertext electronic document for conversion. See abstract) specifying at least one source of content available (i.e. a user is able select one or more hypertext electronic document for conversion. See abstract) via the Internet (i.e. Internet. See abstract), types of data (i.e. display element. See abstract) to be extracted from the at least one source (i.e. a user is able select one or more hypertext electronic document for conversion. See abstract), and fields for structuring extracted items of data (i.e. a user is able select one or more hypertext electronic document for conversion. See abstract);

retrieving content from the at least one source (i.e. the hypertext elements are extracted from. See abstract);

parsing the retrieved content to extract items of data of the types specified by the user (i.e. the selected content types. Col. 15, lines 44-47); and

mapping (i.e. conversion. Col. 15, lines 44-47) the extracted items of data to the fields (i.e. content conversion. Col. 15, lines 33-67) specified by the user (i.e. the selected content types. Col. 15, lines 44-47) so as to transform the extracted items of data into a structured format (i.e. content conversion. Col. 15, lines 33-67).

Under Section 102, a claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in the single prior art reference. (See, e.g., MPEP Section 2131.) As will be shown below, Whitledge fails to

teach each and every element set forth in Applicant's claims 1-63, and therefore fails to establish anticipation of the claimed invention under Section 102.

Whitledge's system enables a user to select one or more display elements from one or more hypertext electronic documents (e.g., Web pages available on the Internet or intranet) and converted into a format suitable for display on a user device based on user conversion preferences (Whitledge, abstract). The selected hypertext elements are extracted and converted using data mining conversion operations (Whitledge, abstract). The data mining conversion operations allow a user to extract only desired display information displayed from a hypertext element and convert the display information into a different format that is more appropriate for the device or environment on which the information will be displayed to the user (e.g., in a format appropriate for display on a cell phone or other hand-held device). Whitledge's solution allows a network device to receive a converted electronic document based on user-conversion preferences, device-conversion preferences, site-specific conversion preferences, or other preferences (Whitledge, col. 17, lines 20-27). The conversion preferences have been previously established (e.g., by a system programmer) and are retrieved from a database as needed when data is to be converted (Whitledge, col. 17, lines 53-65; Fig. 7). Thus, Whitledge contemplates a pre-programmed, mechanical conversion operation that converts content from its original form into another form appropriate for the network of the target device based on previously established conversion preferences (Whitledge, col. 17, lines 20-33).

Applicant's invention, in contrast, does not convert information based on conversion preferences previously established by a system programmer. Instead, Applicant's solution provides the user with additional tools and flexibility in the capture and repurposing of dynamic content for use in a variety of applications. Applicant's solution includes a graphical user interface (GUI) for receiving user commands and data in a graphical (e.g., "point-and-click") fashion (Applicant's specification, paragraph [0079]). This interface can be used to specify not only the elements of information that are to be captured, but also enables the user to aggregate the information into a structured format (Applicant's specification, paragraphs [0149]-[0150]). In addition, Applicant's invention includes methodology for parsing a Web page and creating "fingerprints" of items of dynamic content on the Web page. This fingerprinting involves creating a

feature extraction object (or feature tag) to represent an item of dynamic content (Applicant's specification, paragraphs [0123] - [0124]). Significantly, these feature tags of Applicant's invention can be used to retrieve the same or similar content at a later date, even if content on the page may have changed (Applicant's specification, paragraphs [0068] - [0070]).

Applicant's invention enables a user to capture small objects on a Web page and create a feature tag for the object which represents a Web "fingerprint" of the object. This feature tag may be referenced again to find the object, or other objects having similar attributes, in the future (Applicant's specification, paragraph [0066]). Applicant's solution includes "fuzzy logic" to ensure that targeted content which has been identified and collected can be accessed using the feature extraction tag after a source Web page has been updated with fresh information or graphics. (Applicant's specification, paragraph [0066]). A feature extraction object of Applicant's invention can also be used for identifying similar information objects. This makes it possible to divide and sort content extracted from several pages into groups sharing similar attributes which are included as attributes of the feature extraction object (Applicant's specification, paragraphs [0068], [0157], [0158]). The objects may also be aggregated into a structured format based on these attributes. For example, after selecting items of interest to be extracted, a user may specify that the output be organized into four columns "title", "author", "date" and "text" (Applicant's specification, paragraphs [0156] - [0157]). In response, Applicant's system parses the Web pages to extract the selected items and organizes the information into the four columns based on user input (Applicant's specification, paragraphs [0157] - [0158]; Figs. 11D - 11F). The user may then (optionally) preview the information and make modifications to how it is organized (Applicant's specification paragraphs [0158]; Fig. 11G).

The feature tags of Applicant's invention include attributes of the extracted items and also indicate how the items are to be organized. Thus, Applicant's solution enables a user that is not a programmer to select information to be captured and build an organized model for extracting and organizing the captured information (Applicant's specification, paragraph [0160]). The model that is built can be used a later date to capture and organize information from the selected Web pages even though the particular information

(e.g., headlines, articles or the like on a news Web site) may have changed (Applicant's specification, paragraph [0160]). This provides significant advantages given the rapidly changing nature of dynamic content. The feature tag can be used as an alias pointing to target object(s) of interest which can be used to find the objects of interest even if the Web page(s) including the object(s) has changed (Applicant's specification, paragraphs [0067] - [0070]).

Applicant's claims have been amended in an effort to more clearly articulate these features of Applicant's invention which provide for creation of feature tags which enable subsequent retrieval of items of content having attributes matching the feature extraction tag in a format specified by the user. For example, Applicant's amended claim 1 includes the following claim limitations:

receiving input of a user specifying at least one source of content available via the Internet, types of data to be extracted from said at least one source, and fields for structuring extracted items of data;
retrieving content from said at least one source;
parsing the retrieved content to extract items of data of the types specified by the user;
mapping the extracted items of data to the fields specified by the user so as to transform the extracted items of data into a structured format;
generating a feature tag for each extracted item of data, the feature tag identifying attributes of the item of data and the structured format of the item; and
in response to a subsequent request for an item of data, using the feature tag to obtain the item of data and transform it into the structured format.

(Applicant's claim 1, as amended, emphasis added)

Although Whitledge describes retrieval of content and conversion of content into different formats, Applicant's careful review of the Whitledge reference finds no teaching of creating feature extraction tags identifying attributes of items of content in the manner specified in Applicant's specification and claims. Also, the data mining conversion operations described by Whitledge involve conversion of content based on previously established conversion preferences (e.g., based on characteristics of devices on which the content is to be displayed). Applicant's invention provides a much more flexible solution which enables a user to specify types of information to be collected and how collected information is to be organized for display or reuse. Therefore, as Whitledge does not

teach or suggest all of the claim limitations of Applicant's claims 1-63 it is respectfully submitted that the claims distinguish over this reference and overcome any rejection under Section 102.

Any dependent claims not explicitly discussed are believed to be allowable by virtue of dependency from Applicant's independent claims, as discussed in detail above.

Conclusion

In view of the foregoing remarks and the amendment to the claims, it is believed that all claims are now in condition for allowance. Hence, it is respectfully requested that the application be passed to issue at an early date.

If for any reason the Examiner feels that a telephone conference would in any way expedite prosecution of the subject application, the Examiner is invited to telephone the undersigned at 925 465 0361.

Respectfully submitted,

Date: April 12, 2007

/G. Mack Riddle/

G. Mack Riddle; Reg. No. 55,572
Attorney of Record

925 465 0361
925 465 8143 FAX